The Gene Map of *Homo sapiens*: Status and Prospectus

V.A. McKusick

Division of Medical Genetics, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

HISTORICAL-METHODOLOGICAL INTRODUCTION

The first gene to be mapped to a specific chromosome in man—indeed, in any mammal—was that for color blindness, deduced to be on the X chromosome by E.B. Wilson at Columbia University in 1911. A few dozen other X-linked traits (e.g., hemophilia and Duchenne muscular dystrophy) were identified by characteristic pedigree pattern during the next 57 years before the first assignment of a gene to a specific human autosome: Duffy blood group to chromosome 1 by Donahue and colleagues (1968) at Johns Hopkins University. They made the assignment by finding evidence of linkage between the Duffy locus and a heteromorphism of chromosome 1 that was segregating in a Mendelian manner in Donahue's family. Soon thereafter, haptoglobin was assigned to chromosome 16 by linkage studies in families with inherited, balanced translocations involving 16 (Robson et al. 1969) and in families with a heritable "fragile site" on 16q (Magenis et al. 1970). About the same time, Weiss and Green (1967) showed the feasibility of assigning specific genes to specific chromosomes (or regions of chromosomes) by interspecies somatic cell hybridization (SCH) (e.g., fusion of cells from mouse and man), a method of genetic study called "parasexual" by Pontecorvo and "an alternative to sex" by Haldane.

Between 1951, when the first successful autosomal assignment of a gene was achieved by the linkage method, and 1968, when the fruitful, alternative mapping method of SCH was introduced, nine autosomal linkages in man (seven pairs, one triplet, and one foursome) had been established by family studies, but the specific autosome carrying each of the linkage groups was not known. For the linkage studies in man, special methods of analysis of pedigree data were necessary because it is not possible to design matings as can be done in experimental organisms. The sib-pair method of Penrose, for example, was used by Mohr in 1951 to establish the first autosomal linkage, that between Lutheran blood group and secretor trait. The method of "lods" (log odds) was elaborated by C.A.B. Smith and by Newton E. Morton in the 1940s and 1950s. Computer methods were started about 1960, by Renwick in particular. One of the most widely used programs for linkage analysis, LIPED, was introduced by Ott (1974, 1976, 1985).

Although the potential usefulness of genetic linkage information to clinical medicine (e.g., in the recognition of the carrier and preclinical states of Mendelian disorders) may have been commented on earlier, the earliest, clearest, and most specific statement of use of the linkage principle in prenatal diagnosis is probably that by Edwards in 1956. After it had been pointed out that sexing of the unborn infant is possible by studies of sex chromatin in amniocytes, obtained by amniocentesis, Edwards (1956) suggested that, given close linkage of a testable marker, prenatal diagnosis of genetic disease in the fetus should be possible by study of amniotic material. As the concluding speaker at the Third International Birth Defects Congress at The Hague in 1969, I emphasized mapping the chromosomes of man as a great exploration for the future that was bound to have important rewards in clinical medicine (McKusick 1970). I used the close linkage of G6PD and classic hemophila (Boyer and Graham 1965) as an example of an opportunity for prenatal diagnosis. McCurdy et al. (1971) actually used this approach for hemophila. Others (e.g., Schrott et al. 1973) applied it to the prenatal diagnosis of myotonic dystrophy, using the linkage to secretor.

By SCH, mutually potentiated by family linkage studies (FLS) of the type that assigned *color blindness* and *Duffy*, a veritable explosion of information on the human gene map occurred in the 1970s. By June 1976, at least one gene had been assigned to each of man's 24 chromosomes—the 22 autosomes, the X, and the Y. The ability to identify uniquely each chromosome by its characteristic banding pattern, by methods developed by Caspersson and his colleagues (1970a,b, 1971) and by others (Patil et al. 1971; Schnedl 1971; Seabright 1971; Sumner et al. 1971), has been of fundamental importance to gene mapping. Not only could otherwise very similar chromosomes, such as those of the C group (nos. 6–12 + X), be individually identified by their stripes, even when intermingled with rodent chromosomes in the hybrid cell, but also the composition of chromosomal rearrangements could often be determined and genes could be mapped to specific bands. Banding methods applied to metaphase chromosomes

---

The Appendix cited throughout this paper and found at the back of this volume is a late version of the Human Gene Map newsletter, which has been prepared periodically since 1973. Figures referred to in the text are found there, and the Tables with roman numerals are there as well.

could demonstrate about 400 bands in the total karyotype; high-resolution cytogenetics as developed by Yunis (1976) and others for application to extended chromosomes in prophase or prometaphase could demonstrate more than twice that number.

In the latter half of the 1970s, the methods of molecular genetics were brought to bear on chromosome mapping — both the localization of genes to specific chromosomes or chromosome bands and molecular mapping down to the nucleotide level. Fundamental to progress in molecular genetics was the discovery, in 1970, of site-specific restriction endonucleases, restriction enzymes produced by bacteria to break down foreign DNA. These enzymes became the scalpel for dissecting the human genome. Cloning of DNA segments (genes) in *Escherichia coli* came in 1972 (Watson and Tooze 1981). Southern (1975), then in Edinburgh, devised an elegant hybridization blot method for displaying restriction fragments of DNA, now commonly called the Southern blot, capitalizing on the property of DNA to bind to nitrocellulose paper.

By the technology of recombinant DNA (Watson et al. 1983), libraries of cloned DNA fragments from the entire human genome were prepared, a well-known one being that of Maniatis (Maniatis et al. 1978); these were genomic clones. With reverse transcriptase, complementary, or copy, DNA (cDNA) clones of human genes were prepared from the corresponding messenger RNA (mRNA), first for the human (hemo)globin genes. Restriction maps were prepared of the (hemo)globin gene regions. Recombinant DNA techniques made it almost as easy to identify and isolate specific human genes as it was to electrophorese abnormal hemoglobins or determine the peptide map of hemoglobins ("fingerprinting"). In 1977, two methods of DNA sequencing (Maxam and Gilbert 1977; Sanger et al. 1977) greatly facilitated mapping to the level of individual nucleotides. In 1981, the complete nucleotide sequence of the human mitochondrial chromosome — all 16,569 bp — was published by Sanger's group (Anderson et al. 1981).

The methods of molecular biology added powerfully to the classic methods of FLS and SCH. Probes, either genomic or cDNA, provided by recombinant DNA technology, were used in combination with SCH for mapping by hybridization in solution ("Cot analysis") (e.g., α- and β-globin loci were assigned to 16p and 11p, respectively; Deisseroth et al. 1977, 1978) or by Southern blot analysis of DNA from somatic cell hybrids. A great advantage was that the gene under study did not need to be expressed in the cultured cell. Essentially, any gene for which a probe was available could be mapped.

Molecular genetics also provided new markers for FLS; as a result, such studies have enjoyed a renaissance (White et al. 1985). The new markers were not polymorphisms of the gene product but rather of the genetic material itself, i.e., variations in nucleotide sequence. Kan and Dozy (1978), using *Hpa*I (for *Hemophilus parainfluenzae*) restriction enzyme, discovered the first human DNA polymorphism, located on the 3' ("downstream") side of the β-globin gene. They, as well as Kurnit (1979) and Solomon and Bodmer (1979), pointed out the potentially great usefulness of DNA polymorphisms as markers for linkage studies. Botstein et al. (1980) suggested that restriction-fragment-length polymorphisms[1] (RFLPs; sometimes pronounced "riflips") as revealed by Southern blots could be used for complete mapping of the human genome. The first polymorphism that was demonstrated in an arbitrary or anonymous (function-unknown) segment of DNA was that of Wyman and White (1980), which was subsequently shown (Balazs et al. 1982; DeMartinville et al. 1982) to be situated at the end of the long arm of chromosome 14 between the α-1-antitrypsin locus and the immunoglobin heavy-chain loci. Demonstration of multiple restriction polymorphisms in a segment of the genome to create a haplotype useful in FLS was the work of Kazazian, Orkin, and their colleagues at Johns Hopkins and Harvard, among others, who used the method in the DNA diagnosis of the thalassemias. The clinical usefulness of linkage using RFLPs as markers was dramatically demonstrated by the mapping of Huntington's disease by Gusella et al. (1983), of adult polycystic kidney disease by Reeders et al. (1985), and of cystic fibrosis by several groups in late 1985. In all three of these disorders, possibilities for diagnosis and understanding were opened up, as will be elaborated upon later in this review.

In recent times (mainly since 1980), direct methods of human chromosome mapping have been added to our armamentarium: (1) in situ hybridization of radiolabeled (or immunofluorescence-labeled) DNA sequences ("probes"), generated by recombinant DNA techniques, directly to spreads of chromosomes to identify the site of that piece of DNA ("gene") on the chromosome (Gerhard et al. 1981; Harper et al. 1981; Szabo and Ward 1982); and (2) fluorescence-activated sorting of chromosomes (e.g, Young et al. 1981) followed by application of molecular techniques to determine the gene content of the isolated chromosomes.

Meanwhile, fine mapping of segments of DNA up to lengths of 50–60 kb or more and down to the level of individual nucleotides has been advancing through the application of recombinant DNA technology, restriction endonucleases, and DNA sequencing. The first and perhaps the best example in man is the mapping of the 50-kb segment of the short arm of chromosome 11 that contains five genes for β-globin and the β-like globins of hemoglobins. Such fine-mapping studies revealed an "unexpected complexity of eukaryotic genes" (Watson et al. 1983).

Although the phenotype associated with most chromosomal aberrations is relatively uninformative as to the precise gene content of the chromosomes involved (Lewandowski and Yunis 1977), with the improved res-

---

[1]A.C. Wilson (pers. comm.) objects to the designation restriction-fragment-length polymorphism, introduced by Botstein (1980), because of confusion with one general class of mutants, i.e., length mutants as opposed to substitutions and rearrangements.

olution provided by the banding methods small dele-
tions and other aberrations were found to be associ-
ated with specific phenotypes, especially tumors (e.g.,
retinoblastoma and Wilm's tumor), but also malfor-
mation syndromes such as the Prader-Willi syndrome,
the Langer-Giedion syndrome, the Miller-Dieker lis-
sencephaly syndrome, and the Beckwith-Wiedemann
syndrome. The improved methods for studying the
chromosomes involved in rearrangements and molecu-
lar genetic methods for demonstrating and mapping
oncogenes combined to greatly further the understand-
ing of cancer, including particularly hematologic ma-
lignancies (Mitelmann 1983; Yunis 1983).

The four commingling methodologic streams in the
chromosome-mapping field — linkage, chromosomal,
somatic cell hybridization, and molecular — are mu-
tually potentiating. A combination of methods is often
used, as illustrated by many examples given below. The
data are cumulative; for example, data on the linkage
of two loci collected from successive families (lod scores
are additive), linkage data on the several loci in a stretch
of chromosome, data on physical mapping provided by
somatic cell hybridization, and data on the genetic map
accumulated by linkage studies in families.

The explosion of information on the human gene
map in the last 15 years is reflected in the numbers given
in Table II (Appendix). Eight international workshops
on human gene mapping[2] (HGM-1 through HGM-8)
(Table I, Appendix) have been critical to the collation
and validation of the information on the gene map de-
rived from many different laboratories working with
methods as diverse as lod scores in family linkage anal-

ysis on the one hand and DNA hybridization charac-
teristics on the other.

THE STATUS OF THE HUMAN GENE MAP

Figure 1 (Appendix) presents a pictorial synopsis of
the present status of the human gene map. Three levels
of confidence (confirmed, provisional, or "in limbo")
with which the genes have been assigned are indicated
by different letter styles of the gene symbols. Gene
clusters are indicated by large letters. The Key for Fig-
ure 1 (Appendix) gives not only the definition of the
gene symbols (and synonymous symbols) but also in-
formation on the regional assignment and the method
of assignment (see Appendix for the definition of the
symbols used to designate methods).

As reflected by Figure 1 (Appendix), the chromo-
some that carries each of about 900 structural genes is
known and many of these genes have been fairly pre-
cisely regionalized. This number represents about 47%
of the well-established loci cataloged in Mendelian In-
heritance in Man (McKusick 1986b) and 23% of all loci
cataloged there. (There is an element of circularity in
these figures; increasingly in the last decade, entries
have been created in Mendelian Inheritance in Man for
loci identified by somatic cell genetic or molecular ge-
netic methods, especially if they have been mapped,
even though no allelic variation had been identified.)
The numbers are impressive when viewed in relation to
the rather short period of time that mapping of the
autosomes has been going on. In 1964, when the Cold
Spring Harbor Symposium was last devoted to human
genetics, not a single gene had been assigned to a spe-
cific autosome.

The number of loci that have been mapped is less
than 2% of the 50,000 genes that Homo sapiens is
thought to have as a minimum. (The known density of
genes in the small segment of 11p that contains the β-
globin cluster — 5 genes in 50,000 bp — yields an esti-
mate of about 300,000 genes in all, given that there are
about 3 billion nucleotides[3] in the haploid human ge-
nome. The globin genes are somewhat atypically small,
however [Table 1]. It may not be appropriate, further-

---

[2]The first workshop was organized by Frank Ruddle and held in
New Haven in June 1973. The second, known as the Rotterdam Con-
ference, was organized by Dirk Bootsma and held in The Netherlands
in July 1974. The third, organized by Victor McKusick, was held in
Baltimore in October 1975. The fourth, organized by John Hamerton,
was held in Winnipeg in August 1977. The fifth, organized by Kare
Berg, was held in Oslo in June-July 1981. The seventh, organized by
Robert Sparkes, was held in Los Angeles in August 1983. The eighth,
organized by Albert de la Chapelle, was held in Helsinki in August
1985. The first six were sponsored exclusively by the National Foun-
dation-March of Dimes (now March of Dimes Birth Defects Founda-
tion), which publishes the proceedings as part of its Birth Defects
Original Article Series; the proceedings also appear in Cytogenetics
and Cell Genetics (Table I, Appendix).

The published proceedings of the workshops (Table I, Appendix) are
revealing not only from the point of view of advancing numerology
but also from the standpoint of evolving methodology. At the begin-
ning somatic cell hybridization was the main source of information,
supplemented importantly by the family linkage method. By HGM-5
in 1979, molecular genetic methods were contributing substantially,
mainly in connection with somatic cell hybridization. By HGM-6 in
1981, in situ hybridization was beginning to appear on the scene (e.g,
Harper et al. 1981).

The third edition of Mendelian Inheritance in Man (McKusick 1971)
had a listing of the then-known linkages; a single page sufficed for
the listing of all known autosomal assignments (three) and all known
autosomal and X-linked linkage groups. Accelerating progress in
mapping over the last decade is pictorially displayed in successive edi-
tions of Mendelian Inheritance in Man (McKusick 1986b), starting with
the fourth in 1975, in the review by McKusick and Ruddle in Science
(1977), in the review published in 1980 in Journal of Heredity (Mc-
Kusick 1980) and in four successive versions of the Human Gene Map
Newsletter, published every 12 to 14 months in Clinical Genetics, be-
ginning in December 1982 (McKusick 1982a).

---

[3]Kornberg (1980) gave a value of 2.9 billion bp for the human hap-
loid genome from estimates of the amount of DNA per cell and an
average molecular weight per base pair of 660, to give the conclusion
that 1 picogram (10^{-12} mg) of duplex DNA contains 9.1×10^8 bp. The
DNA in single cells was measured by UV microspectrophotometry of
Feulgen-stained cells (Mirsky and Ris 1951; Leuchtenberger et al. 1954),
a method developed 50 years ago by Caspersson (1936). Mirsky and
Ris (1951) wrote: "In a series of careful measurements by Davison and
Osgood (pers. comm.) on human granulocytes and lymphocytes (from
leukemic blood), the DNA per cell of the former was found to be
6.25×10^{-9} mg and that of the latter 5.84×10^{-9} mg. Our own deter-
mination, on human sperm gave 2.72×10^{-9} mg per cell, approxi-
mately one-half the value for the somatic cells." The values given by
Leuchtenberger et al. (1954) were in the same range. Watson (1976)
gave a somewhat larger estimate of the total number of nucleotides
than did Kornberg (1980), namely 3.3 billion. Because of the differ-
ence in size of the XX and XY sex chromosome pairs, a difference of
about 2% would be expected in the DNA of male and female diploid
cells.

Table 1. The Size of Genes

Gene product	Genomic size (kb)	cDNA (mRNA) (kb)	No. of introns
Small			
α-globin	0.8	0.5	3
β-globin	1.5	0.6	2
Insulin	1.7	0.4	2
Apolipoprotein E	3.6	1.2	3
Parathyroid	4.2	1.0	2
Protein C	11	1.4	7
Medium			
Collagen I			
pro-α-1(I)	18	5	50
pro-α-2(I)	38	5	50
Albumin	25	2.1	14
HMG CoA reductase	25	4.2	19
Adenosine deaminase	32	1.5	11
Factor IX	34	2.8	7
LDL receptor	45	5.5	17
Large			
Phenylalanine hydroxylase	90	2.4	12
Factor VIII	186	9	25

more, to base estimates such as this on the density of genes in gene clusters.)

The variety of genes that have been mapped is as impressive as the numbers and indicates the central role of gene mapping in contemporary biomedical research. Mapped have been genes for enzymes of carbohydrate, lipid, steroid, amino acid, and nucleic acid metabolism; for hemoglobins; for serum proteins such as albumin, haptoglobin, ferritin, C-reactive protein, plasminogen, and orosomucoid; for enzymes of lysosomes, cytosol, mitochondria, and peroxisomes; for cell-surface proteins that function as receptors for hormones, growth factors, complement, viruses, and toxins, or remain with incompletely understood function being demonstrated mainly by immunologic distinctiveness (e.g., some blood groups); for histone and nonhistone chromosomal proteins; for DNA repair enzymes (e.g., DNA polymerase, α and β); for the cytosolic-nuclear receptors for hormones (e.g., the androgen receptor [mutant in the testicular feminization syndrome] and the corticosteroid receptor); for enzymes involved in the synthesis of transfer RNAs (tRNAs); for hormones such as insulin, growth hormone, ACTH, somatomammotropin (placental lactogen), and prolactin; for HLA, complement components, interferons, immunoglobins, and T-cell-antigen receptor, involved in host-defense mechanisms; for carrier proteins (e.g., transferrin), for T-cell "markers" such as T4 and T8; for cytochrome P450 enzymes; for coagulation factors and their inhibitors and activators; for growth factors, such as TCGF and EGF, and their respective receptors; for structural elements of the cell, such as spectrin, actin, myosin, desmin, and tubulin; and for structural proteins of the intracellular matrix, such as the collagens. The genetic determinants for ribosomal and U1 small nuclear RNA, and for one form of tRNA, have been

mapped. More than 40 oncogenes (i.e., human DNA sequences homologous to the oncogenic nucleic acid sequence of mammalian retroviruses such as those of murine, feline, and simian sarcomas) have been assigned to specific chromosomes or chromosome regions. The homeo box genes (e.g., on chromosome 17) are examples of genetic determinants of development. In addition, pathologic phenotypes of which the biochemical basis is not yet known have been mapped (e.g., nail-patella syndrome, forms of congenital cataract and spinocerebellar ataxia, myotonic dystrophy, Huntington's disease, cystic fibrosis, and the Prader-Willi syndrome).

Gene clusters have become evident as a striking feature of the organization of the human genome. Clusters are indicated in Figure 1 (Appendix) by large letters and include the following: the three immunoglobulin clusters (on 2p, 14q, and 22q), the two (hemo) globin clusters (on 11p and 16p), the leukocyte interferon cluster (on 9p), the major histocompatibility complex (on 6p), histone complexes (on 1 and 7), growth hormone–placental lactogen complex (on 17q), the metallothionein cluster (on 16q), the myosin heavy-chain cluster (on 17p), and the β-glycopeptide hormone cluster (on 19). There is a cluster of apolipoprotein genes on 11 and a second on 19. The genes for clotting factors VII and X may be clustered (on 13q). The albumin, α-fetoprotein, and *GC* genes constitute a cluster (on 4q). The genes for arginine vasopressin and oxytocin are clustered on 20q. There is a cluster of carbonic anhydrase genes on chromosome 8. (These groupings are called *clusters* rather than *families* because the latter term is reserved for kindreds of genes that have a common ancestral origin and may or may not be syntenic; the α- and β-globin genes, for example, constitute a gene family.) Gene clustering can lead to combined deficiencies when deletion involves two or more gene loci in the cluster. A form of thalassemia in which neither δ nor β-globin chains are produced is a well-documented example; combined deficiencies of apolipoproteins A1 and C3 (MIM 23455) may be an example.

Presented in Figure 2 (Appendix) is the gene map of the mitochondrial chromosome, which is circular, like a bacterial chromosome. It carries 37 genes in all: 13 for proteins, 22 for tRNAs, and 2 for mitochondrial ribosomal RNAs. The mitochondrion has its own protein-synthesizing machinery. Obviously, most of the structural and enzymatic components of the mitochondrion are coded by genes in the nucleus. Each mitochondrion has 2 to 10 chromosomes. Whereas each nuclear chromosome is present normally in only two copies per cell, the mitochondrial chromosome is present in thousands of copies. Mutations in the mitochondrial chromosome can be expected to lead to disorders with patterns of transmission and other characteristics to some extent different from those of Mendelian disorders. The genetics is expected to be more Galtonian than Mendelian.

The Anatomy of the Human Genome

To this point I have used the customary cartographic metaphor. In reviewing the significance of the information, it may be useful to use an anatomic metaphor (McKusick 1980, 1982b). The linear arrangement of genes on our chromosomes is part of our anatomy. Knowledge of the chromosomal and genic anatomy of *Homo sapiens* has given clinical genetics (and medicine as a whole) a neo-Vesalian basis. A veritable revolution has taken place, dramatically in the practice of clinical genetics, but also generally in medicine — witness what has happened in oncology. Just as *De Humani Corporis Fabrica* of Vesalius (1543) was the basis for the physiology of Harvey (1628) and the pathology of Morgagni (1761), the information on chromosomal and genic anatomy is the foundation of our understanding and management of genetic disease in man.

The anatomic metaphor is useful in examining the significance of the mapping information because it leads naturally to a consideration of the morbid anatomy, the comparative anatomy and evolution, the functional anatomy, the developmental anatomy, and the applied anatomy of the human genome. Most of the contributions to this Symposium address one or several of these aspects. I review only selected aspects of each.

The Morbid Anatomy of the Human Genome

Figure 3 (Appendix) is a pictorial representation of the chromosomal location of mutations "causing" disorders (McKusick 1986a). Beside each chromosome are names of disorders "caused" by mutations located thereon. Many inborn errors of metabolism, such as galactosemia and phenylketonuria, have been mapped by demonstration of the location of the gene for the enzyme deficient in each. In most of these there is evidence at the protein level (and in an increasing number at the gene level, as well) that the mutation involves the structural gene for the enzyme. In other disorders, a nonenzymatic protein is known to be altered in the particular disorder, and it was the mapping of the normal gene that gave information on the location of the disease-producing mutation. Sickle cell anemia is an historic example; a recent one is familial amyloid polyneuropathy, which in several of its forms is "caused" by a mutation of the transthyretin (prealbumin) gene located on chromosome 18.

Other disorders for which the basic defect is not known have been mapped by linkage of the disease phenotype to a genetic marker that is in turn mapped, either a polymorphism of the gene product such as Rh blood group or a polymorphism of DNA (RFLPs). Traditionally, mapping has been practical almost only for autosomal dominants (and X-linked recessives, which in the male behave like dominants). However, with the increased power of the polymorphic DNA markers, recessives can also be mapped, witness cystic fibrosis, even though the heterozygote cannot be identified.

The mapping information developed in recent years has modified our classification of genetic diseases and indicates the importance of the large group of disorders that represent *somatic cell genetic diseases*. All malignant neoplasms fall into this category and some congenital malformations (e.g., aniridia) are demonstrably in that category. The locations of determinants of selected malignancies have been included in Figure 3 (Appendix) for illustrative purposes. Basically, some or most autoimmune diseases may fall into the category of somatic cell genetic diseases.

Mendelian syndromes are usually the consequence of pleiotropism of a single mutant gene. The notion that a genetic syndrome is due to the close linkage of two or more genes, each for a separate component of the syndrome, can usually be rejected as naive. Recent observations of deletions that can be seen in high-resolution karyotypes or deduced from family studies of polymorphic markers indicate that syndromes can indeed result from change in linked genes: the WAGR syndrome (11p) and the Langer-Giedion syndrome (8q) are cases in point. Although the evidence is not ironclad, in each it seems that separate components of the syndrome may occur alone. Combined deficiency of C6 and C8 (which are known to be closely linked, although the chromosomal location is not known) and of apolipoproteins A-I and C-III (close together on 11q) are cases in point. See also chronic granulomatous disease (CGD) with and without Xk deficiency and X-linked adrenal hypoplasia with and without GK deficiency.

In Figure 3 (Appendix), allelic disorders, which may be so different in phenotype as to suggest mutation in different genes, are indicated by enclosure in a box. The diversity of the phenotypes caused by mutations in the β-globin gene on 11p is a classic. Conversely, the same phenotype can, of course, be caused by mutation in different genes ("genetic heterogeneity"). Type VII Ehlers-Danlos syndrome can be caused by mutation in either the α-1 chain (Cole et al. 1986) of type I procollagen (coded by 17q) or its α-2 chain (Steinmann et al. 1980) (coded by 7q) — and there may be yet another form of Ehlers-Danlos syndrome type VII determined by mutation, not in a procollagen gene, but in the gene (not yet mapped) for the procollagen peptidase that cleaves the amino-peptide from the procollagen molecule (Lichtenstein et al. 1973). The last has been demonstrated in certain domestic animals though not yet in humans.

Some of the entities indicated in Figure 3 (Appendix) are "nondiseases"; they turn up as abnormal test values in laboratory studies and are important to know about to avoid confusion with diseases. These "nondiseases" include inborn variants of metabolism such as cystathioninuria (on chromosome 16) and pentosuria, an inborn nondisease of metabolism that has not yet been mapped. They also include abnormalities of bind-

ing by albumin, giving high levels of thyroxine or zinc without clinical evidence of intoxication.

Three entries in Figure 3 (Appendix) are infectious diseases for which the role of a single locus in susceptibility or resistance has been identified. The Duffy null gene (on chromosome 1) gives resistance to vivax malaria. As far as known, all humans are susceptible to diphtheria and poliomyelitis (Miller et al. 1974) by reason of the products of genes located on chromosomes 5 and 19, respectively. Vitamin C deficiency is a universal inborn error of metabolism in *Homo sapiens*; where the mutation is in the human genome will be known when the gene for *L*-gulonolactone oxidase is mapped (by now this gene might be only a pseudogene relic, if present at all).

Comparative Anatomy and Evolution of the Human Genome

Footprints indicating the role of gene duplication in its evolution are seen throughout the human genome—in the gene clusters and families and even in the internal structure of genes.

There is some correspondence between exons and domains of proteins. It was suggested by Gilbert (1982) that exon shuffling is a mechanism of evolution whereby exons from various sources are combined to fashion a protein of optimal characteristics for a given function. A possible rationale for introns (intervening sequences) is the opportunity they afford for recombination without disruption of the coding segments (exons).

Because of the considerable similarity in banding pattern of the chromosomes of apes and man (Yunis et al. 1980), it is not surprising that many genes that are known to be syntenic in man have been found to be syntenic in other higher primates (Lalley and McKusick 1985) when appropriate somatic cell hybridization or in situ hybridization studies are done. Furthermore, in the other primates, homologous loci have usually been found to be carried by the chromosome judged by banding pattern to be homologous. The degree of homology of synteny between mouse and man came, however, as a considerable surprise. Ohno's law (Ohno 1973), which predicts identity of the genic content of the X chromosome in all mammals, is a special case. Except perhaps for a few loci at the tip of the short arm of the X chromosome that escape lyonization, X-linkage can be expected to be conserved in all mammals. The ill effects of loss of dosage compensation would be expected to prevent movement of most genes from the X chromosome to an autosome. Because of Ohno's law, X-linked diseases in mice and other mammals (e.g., hereditary hypophosphatemia and testicular feminization) are convincing models of human X-linked diseases.

Comparative mapping has been aided greatly by molecular genetic methods. Whereas criteria for homology of the gene product such as immunologic cross-reactivity and similarities of substrate specificity were previously used, homology can be tested directly by using the same DNA probes in hybridization studies in various species.

Most would not have predicted the degree of autosomal homology of synteny that has been found between such distant relatives as man and mouse (Buckle et al. 1984; Nadeau and Taylor 1984). In the tabulation made at HGM-8 (Lalley and McKusick 1985), all the human autosomes except chromosome 13 are shown to have at least 2 loci that are also syntenic in the mouse. That gap will be filled, perhaps, when the chromosomal locations of genes *F7, F10, COL4A1, COL4A2*, and others on human 13 are known in the mouse. Human chromosome 17 has 8 loci that are all on mouse chromosome 11; the short arm of human chromosome 6 has at least 10 loci (counting all *HLA* loci as one) that are on mouse chromosome 17. Some human chromosomes bear homology in genic content to two or three mouse chromosomes. Thus, chromosome 1 of man has a distal 1p region with 6 loci homologous to loci on mouse 4, a proximal 1p region with 6 loci homologous to loci on mouse 3, and a 1q region with 4 loci homologous to loci on mouse 1. It is useful to consult such a table (Lalley and McKusick 1985) when a given locus has been mapped in mouse or other subhuman species for a guess as to where the human locus may be situated. Buckle et al. (1984) published an ingenious grid that indicates at a glance the synteny homologies between man and mouse.

An example of predicting human chromosomal assignment from findings in the mouse is the following: Because the genes coding for aminoacylase (ACY1) and for β-galactosidase-1 (GLB1) are on chromosome 3 in man and chromosome 9 in the mouse and since the structural gene for transferrin (TF) is closely linked to *ACY1* and *GLB1* in the mouse, Naylor et al. (1980) suggested that the human transferrin gene might be on chromosome 3. This was subsequently shown to be the case (Huerre et al. 1984; Yang et al. 1984). The homology did not extend to close linkage, however; in the human *TF* is on 3q, whereas *GLB1* and *ACY1* are on 3p.

An ancient tetraploidization, partial or complete, has been suggested (Comings 1972) by morphologic similarities, for example, of human chromosomes 11 and 12 and of chromosomes 21 and 22. The genic content of 11 and 12 gives some support to this idea; *LDHA* and *LDHB*, and the Harvey and Kirsten *ras* proto-oncogenes, are on 11p and 12p, respectively. *IGF1* is on 12q and *IGF2* is on 11p. Mouse chromosome 16 carries several loci that are on human chromosome 21, which is trisomic in the Down syndrome; but mouse chromosome 16 also carries the supergene for the λ light chain of immunoglobulin, which in man is on chromosome 22 (Lalley and McKusick 1985). (Other loci on human chromosome 22 are found on chromosome 15 in the mouse [Lalley and McKusick 1985].)

The coding portions (exons) of the two γ-globin genes (in the HBBC on 11p) differ in only a single codon, number 135, resulting in either alanine or glycine

as the 135th amino acid. This reflects not-unexpected gene divergence after duplication, and indeed more difference might be anticipated. The finding of the same restriction polymorphism in noncoding intervening segments (introns) of these two linked genes (Jeffreys 1979) and the same nucleotide sequence (Slightom et al. 1980) may be explained by gene conversion or correction. A similar process has probably operated to preserve identity or close similarity of the two α-globin genes as well as of members of other gene clusters. Proximity of the genes involved is a necessary condition for gene conversion.

Functional Anatomy of the Human Genome

A considerable amount of information can be summarized in the following seven generalizations.

1. Although clustering of genes of similar function and common evolutional origin is a frequent finding, there is *no chromosomal aggregation of genes coding for structure and function of particular organs, such as the eye, heart, or kidney, or particular subcellular organelles, such as lysosomes or mitochondria.*

2. *The structural genes for enzymes catalyzing successive steps in a particular metabolic pathway are usually not syntenic.* The genes for at least three enzymes of galactose metabolism (GALE, GALK, and GALT), for five enzymes of the urea cycle (ARG1, ASL, ASS, CPS1, and OTC), and for eight enzymes of the tricarboxylic acid cycle (ACO1 and 2, IDH1 and 2, FH, MDH1 and 2, and CS) are known to be on separate chromosomes. On the other hand, the genes for at least four enzymes involved in glycolysis are all on chromosome 12 (TPI, CAPD, ENO1, and LDHB and probably LDHC). Glucose dehydrogenase (GDH) and 6-phosphogluconate dehydrogenase (PGD), enzymes that catalyze successive steps in the phosphogluconate pathway, are coded by linked genes on the short arm of chromosome 1, but the genes for two other enzymes of this pathway (G6PD and GAPD) are on other chromsomes. There may be a functional relationship between HPRT and PRPS, enzymes coded by genes rather closely situated on Xq. Of the nine enzymes of the purine ribonucleotide biosynthetic pathway, two are coded by genes on each of two different chromosomes: PGFT and PFGS by chromosome 14 and PAIS and PRGS by chromosome 21. OPRT and ODC, enzymes involved in successive steps of the pyrimidine synthesis pathway, are both coded by chromosome 3 and both are mutant in most cases of hereditary orotic aciduria (only ODC is mutant in a single known case). All these are not exceptions, however, because in the case of each set the enzyme activities are properties of a single multifunctional protein (D. Patterson, pers. comm.). The single bifunctional enzyme deficient in orotic aciduria is called uridylmonophosphate synthase (Patterson et al. 1983). A trifunctional enzyme that has been conserved in *Drosophila*, birds, and mammals is coded by chromosome 21. Mutants of each of the three enzymatic functions individually are known in human cells and a double mutant of the PAIS and PRGS functions are known in Chinese hamster ovary cells. It is of interest that, in the case of both the PGFT/PFGS and the PAIS/PRGS multifunctional enzymes, the reactions catalyzed are not contiguous in the metabolic chain.

There are several other examples of enzymatic functions at steps in the same metabolic processes being subserved by a single multifunctional molecule. The advantage of this arrangement is that equimolar synthesis of the two enzymatic entities is guaranteed. Linkage of the structural genes of thymidine kinase and galactokinase (on human chromosome 17) has been conserved over long evolutionary time. Coordinate function may be responsible for this (Schoen et al. 1984). It may have functional significance that *PFKP* and *HK1*, the genes for enzymes at the primary and secondary control points in the glycolytic pathway, are both on 10p.

3. *The genes determining the different subunits of a heteromeric protein are usually not syntenic.* The α and β chains of adult hemoglobin (Hb A), determined by genes on chromosome 16p and 11p, respectively, are cases in point. Other examples are shown in Table 2. The class II HLA proteins (e.g., HLA-DR), of which the α and β chains are determined by separate loci, both of which are in the major histocompatibility complex (MHC) on 6p, represent an exception to this rule of nonsynteny. Another exception is fibrinogen, of which the α, β, and γ chains are coded by chromosome 4; indeed, the genes are in the same order as the polypeptides in the fibrinogen molecules—γ-α-β (Aschbacher et al. 1985; Kant et al. 1985). Insulin and haptoglobin do not represent exceptions; in both cases the two chains are coded by a single gene with posttranslational cleavage of the proprotein into two. These are examples of proteins that are coded by a single gene but in the mature or active form consist of two subunits held together by disulfide bonds; activated PLAT and the α-γ complex of C8 are other such instances.

Table 2. Nonsynteny of Genes Coding for Subunits of Heteromeric Proteins

Coagulation factor XIII	A,B	6p, not 6p
Creatine kinase	B,M	14q,19q
Collagen, type I	α1, α2	17q,7q
Ferritin	H,L	11,19q
Glycopeptide hormones		
chorionic gonadotropin	α,β	6q,19q
follicle-stimulating hormone	α,β	6q,11p
luteinizing hormone	α,β	6q,19q
thyroid-stimulating hormone	α,β	6q,1p
Hemoglobin	α,β	16p,11p
Hexosaminidase	α,β	15q,5q
HLA-A,-B,-C	H,L	6p,15q
Immunoglobulins	H,L	14q;2q,22q
Lactate dehydrogenase	A,B	11p,12p
Phosphofructokinase, red cell	L,M	21q,1q
Platelet-derived growth factor	A,B	7,22
Protein kinase C'	α,β,γ	17,16,19
T-cell antigen receptor	α,β,σ,γ,ε	14q,7q,7p,11q,?

For information on mapping and other genetic aspects, see McKusick (1986b).

It must be asked whether the nonsynteny of heteromers is more than what one would expect given a random distribution as the general rule. It would appear that there is a true repulsion of the genes for the several heteromers of a protein, especially when one considers that many or even most probably originated by duplication of a common ancestral gene.

4. Whereas most heteromeric proteins are compounded of polypeptides coded by genes on different chromosomes, *some genes code for more than one polypeptide*. As just mentioned, the α and β chains of haptoglobin are coded by a single gene (on 16q) and the A and B chains of insulin by a single gene (located near the distal end of 11p). A striking example of multiple peptides from a single gene is proopiomelanocortin (on 2p); ACTH, β-endorphin, and β-melanotropin are three of the some seven peptides derived from the same gene. Tissue-specific alternative splicing of the primary RNA transcript is a method by which a single gene can code for proteins specifically suited to the differentiated function of different cell types. The calcitonin gene (on 11p) in the parafollicular cells of the thyroid codes for calcitonin but in the hypothalamus codes for calcitonin gene–related peptide (CGRP), which is read off the same primary transcript. The mechanism of this differential function is unknown.

5. *The genes determining the cytoplasmic and mitochondrial forms of a given enzyme are not syntenic.* The cytosolic (cytoplasmic or soluble) and mitochondrial forms, referred to as -1 and -2, respectively, of the following enzymes are determined by different chromosomes: ACO (9 and 22), ALDH (9 and 12), GOT (10 and 16), IDH (2 and 15), MDH (2 and 7), SOD (21 and 6), and TK (17 and 16). I know of no true exception to this rule of nonsynteny. Adenylate kinase exists in a cytosolic form (AK1) determined by a gene on 9p and in two mitochondrial forms (AK2 and AK3) determined by genes on 1p and 9q, respectively. Since the genes for AK1 and AK3 are on different arms of a long chromosome, this is probably not an exception to the rule. Similarly, ME1 (on 6q) and ME2 (on distal 6p) probably do not represent an exception. That both cytosolic and mitochondrial fumarate hydratase (FH) are determined by chromosome 1 is also not an exception: 1q carries a single structural gene for FH; posttranslational modification accounts for the electrophoretic differences in the two isozymes (Edwards and Hopkinson 1979). These observations of nonsyntenic genetic determination of cytosolic and mitochondrial isozymes are consistent with a symbiont origin of the mitochondria, with a shift of most of the mitochondrial genes from the mitochondrial chromosome to nuclear chromosomes in a random manner, and with no more homology between the mitochondrial gene and the nuclear gene for each pair of isozymes than might be expected on the basis of a very ancient origin of both from a common ancestral gene.

6. *An appreciable portion of the genome consists of functionless (unexpressed) pseudogenes*, which show similarities in nucleotide sequence to functional genes.

Pseudogenes have lost critical elements necessary for transcription. Because of sequence homology to functional genes, however, they are recognized by the same DNA probes. Pseudogenes may be closely situated to the structural genes of which they are imperfect replicas (e.g., the pseudogenes in the α- and β-globin gene clusters) or may be far removed and present in many copies, as in the case of the pseudogenes of argininosuccinate synthetase (Su et al. 1984). The lack of introns in pseudogenes suggests that they are "processed genes," that is, they originated by integration of reverse transcripts of mRNA. The differentiation of functional genes from pseudogenes is aided by somatic cell hybridization; for example, the functional gene for argininosuccinate synthetase is the one demonstrated on chromosome 9 by ISH because the enzymatic function maps to chromosome 9 by SCH. (HGM-8 tentatively recognized another category of gene called "like." These are identified by in situ and other molecular hybridization methods under conditions of low stringency. The functional status of these or their relation to pseudogenes is unknown; hence the noncommittal designation.)

7. *The structural gene for a receptor and that for its ligand are usually not on the same chromosome.* Both transferrin and the transferrin receptor are coded by 3q; however, the genes are rather far apart in the 3q21 and 3q26.2 bands, respectively, and the transferrin receptor bears no sequence homology to transferrin (McClelland et al. 1984). The LDL receptor is coded by chromosome 19, as is also one of its ligands, apolipoprotein E, but the genes are on 19p and 19q, respectively. *CSF1* and *CSF1R* may be in the same band on 5q. The exceptions are, however, more numerous than the nonexceptions (see Table 3); for example, epidermal growth factor (EGF) is coded by chromosome 4, whereas the gene for its receptor is on chromosome 7.

Functional significance can, perhaps, be attached to the clustering of the various components of the MHC on 6p. These genes include not only the *HLA* loci of classes I and II, but also the determinants of certain components of the complement and alternative path-

Table 3. Sometimes the Genes for a Receptor and Its Ligand(s) Are Syntenic but Usually Not

Examples of synteny			
CSF1R(FMS)	5q	CSF1	5q[a]
LDLR	19p	APOE	19q
TFR	3q26.2	TF	3q21
Examples of nonsynteny			
EGFR	7p	EGF	4q
IFNAR	21q	IFNA	9p
IFNBR	21q	IFNB	9p
IFNGR	18	IFNG	12q
IGF1R	15q	IGF1	12q
INSR	19p	INS	11p
NGFR	17q	NGFB	1p
PDGFR	5q	PDGFA,B	7,22

[a]Only known example of mapping to the same band.

ways. The convertase involved in activation of C3 (gene assigned to chromosome 19) is a bimolecular complex of C4 and C2, both of which are coded by genes closely linked to *HLA-B* on chromosome 6. Properdin factor B (BF), which serves a similar role (of activating C3) in the alternative pathway, is also closely linked to *HLA-B*. (The genes for C6 and C7, linked in the dog and the marmoset, are also closely linked in man, as indicated by restriction enzyme mapping studies and by observation of combined deficiency. The genes are not on 6p, however.) No functional significance is evident for the location within the MHC of genes for 21-hydroxylase deficiency (*CAH1*) and hemochromatosis (*HFE*).

The close situation on 11p of the genes for parathyroid hormone and calcitonin (the yin and yang of calcium homeostasis) is probably happenstance and of no evolutionary or functional significance. The dissimilarity in sequence of the genes (and the peptides they determine) rules against their origin from a common ancestral gene. Possibly in favor of a functional significance of their close situation is the fact that both are on mouse chromosome 7 (P.A. Lalley, pers. comm.), which carries other genes, that are on human 11p, such as the genes for insulin, β-globin, LDH-A, and the β subunit of follicle-stimulating hormone, as well as the Harvey-*ras* oncogene.

The Developmental Anatomy of the Human Genome

The linear orientation of the cluster of β-globin genes (HBBC) on 11p appears to have ontogenetic significance. During development, the ϵ gene (at the 5' end of the 50-kb segment) is active during the embryonic period. Later, switch occurs to the two γ genes, which are next downstream from the ϵ gene and are active during the fetal period, and then to the δ and β genes, which are active during postnatal life. The gene for α-fetoprotein, the fetal equivalent of serum albumin and a protein of diagnostic usefulness to the oncologist and medical geneticist, is closely linked to albumin on 4q. Curiously, in the mouse where the two loci are also closely linked, the postnatally active albumin gene is upstream from (i.e., on the 5' side of) the gene for the fetal counterpart, a situation opposite to that for the non-α-globin genes of mouse and man. It turns out, however, that the gene-switching paradigm of globin ontogeny is not precisely applicable to the *AFP-ALB* system; the albumin gene is active throughout development, whereas the *AFP* gene, active in embryonic and fetal stages, is largely switched off in the postnatal period.

The genetics of differentiation, and specifically the significance of the anatomy of the human genome to morphologic development and differentiated function, are largely unknown. Among the many aspects of human biology that have been illuminated by the study of hemoglobins, this is one: the nondeletion (or heterocellular) type of hereditary persistence of fetal hemoglobin appears to result from mutation in a regulator for switch from γ- to β-globin synthesis. Tight (Old et al. 1982) and loose (Gianni et al. 1983) linkage of the mutant regulator(s) to the non-α-globin cluster has been found.

The ontogeny of the immunoglobulin-producing lymphocyte appears to be related to the anatomic orientation of the several components of the three immunoglobulin gene clusters, those for the heavy chain (on chromosome 14) and for the $\varkappa$ (on chromosome 2) and λ (on chromosome 22) light chains. Generation of diversity in antibodies through somatic gene rearrangements is dependent on close linkage of the *V, D, J,* and *C* genes that make up the immunoglobulin gene clusters. The developmental significance of the anatomy of the immunoglobulin genes is seen also in the case of the different genes for the *C*, or constant, part of the immunoglobulin heavy chain. Splicing of various *V, D,* and *J* genes provides diversity; the constant region gene that is closest to *D*, or diversity-generating, part of the complex is the gene activated first. Thus, production of IgM occurs early in the immune response and the switch to one or another of the constant region genes for production of IgD, IgG, IgE, and IgA ("class switch") takes place later. (Rather than representing a cluster of genes, each of the immunoglobulin-determining segments of DNA can be viewed as a single supergene in which the diversity-generating portions and those coding for the constant regions are exons.)

The rearrangements of the T-cell antigen receptor genes are another example of developmental significance of genomic anatomy. Like the immunoglobulins, the T-cell antigen receptor consists of two polypeptide chains. The α and β chains of the T-cell receptor are coded by chromosomes 14 and 7, respectively. The genes, symbolized by *TCRA* and *TCRB*, are a cluster of genes (supergene) with *V, D, J,* and *C* genes coding for constant and variable domains of the T-cell receptor molecules. The maturation of the T-cell in the thymus involves clonal rearrangement within the gene cluster to bring the constant-coding gene into contiguity with one of the variable region genes. Another *TCR* gene called γ (*TCRG*) is also situated on chromosome 7 but is on 7p, whereas *TCRB* is on 7q.

The function of the T-cell antigen receptor is to recognize antigens in combination with the individual's own MHC proteins. In the thymus, precursor lymphocytes destined to become T cells undergo a period of "thymic education." Immature T cells that respond to one or a small group of MHC proteins are allowed to propagate and continue their differentiation. The α gene (on chromosome 14) is little expressed in the immature T cell, whereas the γ and β genes (on 7p and 7q, respectively) produce a large amount of protein. Tonegawa's group (Tonegawa 1985) suggested that a switch occurs from γ-β to α-β with maturation of T cells. Obviously, since *TCRG* and *TCRA* are on separate chromosomes, the switch from γ to α is independent of anatomic proximity, unlike the β switch in hemoglobin synthesis.

The Applied Anatomy of the Human Genome

The reason that great interest has accompanied the mapping of Huntington's disease, cystic fibrosis, adult polycystic kidney disease, myotonic dystrophy, Duchenne muscular dystrophy, and other disorders is at least twofold. All of these disorders are the result of presently unknown, basic defects. For that reason, no thoroughly satisfactory diagnostic test or therapy can be designed on the basis of fundamental defect. Mapping information opens the possibility for diagnosis on the basis of linkage principle. Furthermore, it holds out hopes of determining the basic gene defect by "reverse genetics" ("chromosome walking") and using that information to devise tests for prenatal, preclinical, and carrier diagnosis. These tests may take the form of direct testing of DNA for a gene defect by a process one might call "biopsy of the human genome." Information on the basic defect may help plan methods for ameliorating the disorder even though gene therapy is not possible in the near future.

In addition to "chromosome walking," long-range mapping, and other methods for pinpointing the defect in DNA, determining the basic defect can also follow the "candidate gene" strategy. Given a protein that is a plausible candidate for the site of the basic defect, one can ask: Do the disease and the molecule map to the same area? Does the disease show linkage with a RFLP related to the cloned gene? In persons with the given disorder, is there structural abnormality of the gene for the given protein?

The Rh-linked form of elliptocytosis is probably due to mutation in the gene for protein 4.1 of the red cell membrane because they map to the same area of 1p. On the other hand, Wilson's disease (on 13q) cannot be due to mutation in the structural gene for ceruloplasmin (on 3q); nor can hemochromatosis (on 6p) be due to mutation in the structural gene for transferrin (on 3q), transferrin receptor (on 3q), ferritin light chain (on 19q), or ferritin heavy chain (on 11). (Perhaps those walking 6p in the region of the class I MHC genes will stumble on the hemochromatosis gene, which appears to be near *HLA-A* on its centromeric side.)

Even though no gross abnormality such as deletion or rearrangement can be demonstrated in the *COL1A2* gene on chromosome 7 in these cases, linkage between a *COL1A2* RFLP and osteogenesis imperfecta type IV strongly suggests that the causative mutation is in that gene (Falk et al. 1985; Grobler-Rabie et al. 1985). Phillips et al. (1981) could show that mutation in the growth hormone gene was responsible for pituitary dwarfism in some cases in which Southern blot analysis showed it to be deleted.

PROSPECTUS

Complete mapping of the human genome and complete sequencing are one and the same thing. They must go hand in hand. Nucleotide sequencing will be done within, out from, and between genes that have been localized as precisely as possible in relation to recognized landmarks, the chromosome bands, and in relation to neighboring genes. A RFLP map has properties like both a map of expressed genes and a complete nucleotide sequence. A reasonably detailed RFLP map will have great usefulness in both the mapping of expressed genes and the complete sequencing.

The potential usefulness of complete mapping/sequencing has been emphasized by some interested in birth defects (McKusick 1970) and by others interested in cancer (Dulbecco 1986). The usefulness is, in only a relatively restricted manner, indicated in the earlier section on the Applied Anatomy of the Human Genome. Great value is seen in the understanding of multifactorial disorders, a category into which most cancers fall. Dissecting out the role of individual genetic factors in disorders such as essential hypertension, atherosclerosis, mental illness, and common forms of congenital malformations promises to be valuable in the identification of unusual vulnerability and in planning preventive strategies. Characteristics that are patently genetic but presently not analyzable, such as special talents (e.g., musical and mathematical) and morphologic traits (e.g., facial characteristics, eye color, and attached/unattached earlobes), might be studied successfully, given a detailed RFLP map, for example.

The task of complete mapping/sequencing will require new techniques and improvements in existing ones. A large need, which is addressed in some of the papers in this Symposium volume, is for methods to bridge the gap between the resolution that is achieved by restriction mapping and nucleotide sequencing of overlapping cosmid clones (up to 50 or 100 kb) and that achieved with chromosome banding and linkage analysis (down to 1000 kb at best). Pulsed-field gel electrophoresis (Schwartz and Cantor 1984; Smith and Cantor 1986) is one method that can help bridge the gap. The rallying cry is for completion of total mapping/sequencing by the year 2000 or before. The magnitude of the task is indicated by the fact that the human haploid genome contains about 3.0 billion nucleotides.

Complete mapping/sequencing of the mitochondrial chromosome has been achieved. This is the goal to which mapping of the nuclear genome aspires. Anderson et al. (1981) filled three closely printed pages of *Nature* with the sequence of the mitochondrial chromosome. To print the sequence of the haploid nuclear genome of a single person in a similar manner (the nuclear genome is about 200,000 times larger) would require the equivalent of about 13 sets of the *Encyclopedia Britannica*. To print also the heterozygous variation in that individual and add the enormous range of variation between individuals will require the utmost in computer facilities. Obviously there is a large library task here and a large problem in recovering and reading the information as well as problems in the creation of indexes and concordances and devising methods for recognizing pattern similarities.

REFERENCES

Anderson, S., A.T. Bankier, B.G. Barrell, M.H.L. deBrujin, A.R. Coulson, J. Droin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schrier, A.J.H. Smith, R. Staden, and I.G. Young. 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290**: 457.

Aschblacher, A., K. Buetow, D. Chung, S. Walsh, and J. Murray. 1985. Linkage disequilibrium of RFLP's associated with α, β, and γ fibrinogen predict gene order on chromosome 4. *Am. J. Hum. Genet.* **37**: A186 (Abstr.).

Balazs, I., M. Purrello, P. Rubinstein, A. Alhadeff, and M. Siniscalco. 1982. Highly polymorphic DNA site D14S1 maps to the region of Burkitt lymphoma translocation and is closely linked to the heavy chain γ-1 immunoglobulin locus. *Proc. Natl. Acad. Sci.* **79**: 7395.

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**: 314.

Boyer, S.H., and J.B. Graham. 1965. Linkage between the X chromosome loci for glucose-6-phosphate dehydrogenase electrophoretic variation and hemophila A. *Am. J. Hum. Genet.* **17**: 320.

Buckle, V.J., J.H. Edwards, E.P. Evans, J.A. Jonasson, M.F. Lyon, J. Peters, A.G. Searle, and N.S. Wedd. 1984. Chromosome maps of man and mouse. II. *Clin. Genet.* **26**: 1.

Caspersson, T. 1936. Ueber den chemischen Aufbau des Strukturen des Zellkernes. *Skand. Arch. Physiol.* (suppl. 8) **73**: 1.

Caspersson, T., C. Lamakka, and L. Zech. 1971. Fluorescent banding. *Hereditas* **67**: 89.

Caspersson, T., L. Zech, and C. Johansson. 1970a. Differential banding of alkylating fluorochromes in human chromosomes. *Exp. Cell Res.* **60**: 315.

Caspersson, T., L. Zech, C. Johansson, and E.J. Modest. 1970b. Quinocrine mustard fluoroscent banding. *Chromosoma* **30**: 215.

Cole, W.G., D. Chan, G.W. Chamber, I.D. Walker, and J.F. Bateman. 1986. Deletion of 24 amino acids from the pro-α-1(I) chain of type I procollagen in a patient with the Ehlers-Danlos syndrome type VII. *J. Biol. Chem.* **261**: 5496.

Comings, D.E. 1972. Evidence for ancient tetraploidy and conservation of linkage groups in mammalian chromosomes. *Nature* **238**: 455.

Deisseroth, A., A. Neinhuis, J. Lawrence, R. Giles, P. Turner, and F. Ruddle. 1978. Chromosomal localization of human β-globin gene on human chromosome 11 in somatic cell hybrids. *Proc. Natl. Acad. Sci.* **75**: 1456.

Deisseroth, A., A. Nienhuis, P. Turner, R. Velez, W.F. Anderson, F. Ruddle, J. Lawrence, R. Creagen, and R. Kucherlapati. 1977. Localization of the human α-globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridization assay. *Cell* **12**: 205.

DeMartinville, B., A.R. Wyman, R. White, and U. Francke. 1982. Assignment of the first random restriction fragment length polymorphism (RFLP) locus (D14S1) to a region of human chromosome 14. *Am. J. Hum. Genet.* **34**: 216.

Donahue, R.P., W.B. Bias, J.H. Renwick, and V.A. McKusick. 1968. Probable assignment of the Duffy blood group locus to chromosome 1 in man. *Proc. Natl. Acad. Sci.* **61**: 949.

Dulbecco, R. 1986. A turning point in cancer research: Sequencing the human genome. *Science* **231**: 1055.

Edwards, J.H. 1956. Antenatal detection of hereditary disorders (letter). *Lancet* **I**: 579.

Edwards, J.H. and D.A. Hopkinson. 1979. The genetic determination of fumarase isozymes in human tissues. *Ann. Hum. Genet.* **42**: 303.

Falk, C.T., R.C. Schwartz, F. Ramirez, and P. Tsipouras. 1985. Use of molecular haplotypes specific for the human pro-α-2(I) collagen gene in linkage analysis of the mild autosomal dominant forms of osteogenesis impertecta. *Am. J. Hum. Genet.* **38**: 269.

Gerhard, D.S., E.S. Kawasaki, F.C. Bancroft, and P. Szabo. 1981. Localization of a unique gene by direct hybridization in situ. *Proc. Natl. Acad. Sci.* **78**: 3755.

Gianni, A.M., M. Bregni, M.D. Cappellini, G. Giorelli, R. Taramelli, B. Giglioni, P. Comi, and S. Ottolenghi. 1983. A gene controlling fetal hemoglobin expression in adults is not linked to the non-α-globin cluster. *EMBO J.* **2**: 921.

Gilbert, W. 1982. DNA sequencing and gene structure (Nobel lecture). *Science* **214**: 1305.

Grobler-Rabie, A.F., G. Wallis, D.K. Brebner, P. Beighton, A.J. Bester, and C.G. Mathew. 1985. Detection of a high frequency *Rsa*I polymorphism in the human pro-α-2(I) collagen gene which is linked to an autosomal dominant form of osteogenesis imperfecta. *EMBO J.* **4**: 1745.

Gusella, J.F., N.S. Wexler, P.M. Conneally, S.L. Naylor, M.A. Anderson, R.E. Tanzi, P.C. Watkind, K. Ottina, M.R. Wallace, A.Y. Sakaguchi, A.M. Young, I. Shoulson, E. Bonilla, and J.B. Martin. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**: 234.

Harper, M.E., A. Ullrich, and G.F. Saunders. 1981. Localization of the human insulin gene to the distal end of the short arm of chromosome 11. *Proc. Natl. Acad. Sci.* **78**: 4458.

Huerre, C., G. Uzan, K.H. Grzeschik, D. Weil, M. Levin, M.-C. Hors-Cayla, J. Boue, A. Kahn, and C. Junien. 1984. The structural gene for transferrin (TF) maps to 3q21-3qter. *Ann. Genet.* **27**: 5.

Jeffreys, A.J. 1979. DNA sequence variants in the Gγ-, Aγ-, δ- and β-globin genes of man. *Cell* **18**: 1.

Kan, Y.W. and A.M. Dozy. 1978. Polymorphisms of DNA sequence adjacent to human β-globin structural gene: Relationship to sickle mutation. *Proc. Natl. Acad. Sci.* **75**: 5631.

Kant, J.A., A.J. Fornace, Jr., D. Saxe, M.I. Simon, O.W. McBride, and G.R. Crabtree. 1985. Evolution and organization of the fibrinogen locus on chromosome 4: Gene duplication accompanied by transposition and inversion. *Proc. Natl. Acad. Sci.* **82**: 2344.

Kornberg, A. 1980. *DNA replication*, p. 19. W.H. Freeman, San Francisco.

Kurnit, D.E. 1979. Evolution of sickle variant gene. (Letter). *Lancet* **I**: 104.

Lalley, P.A. and V.A. McKusick. 1985. Report of the committee on comparative mapping (HGM8). *Cytogenet. Cell Genet.* **40**: 498.

Leuchtenberger, C., R. Leuchtenberger, and A.M. Davis. 1954. A microspectrophotometric study of the desoxyribose nucleic acid (DNA) content of cells of normal and malignant human tissues. *Am. J. Pathol.* **30**: 65.

Lewandowski, R.C. and J.J. Yunis. 1977. Phenotypic mapping in man. In *New chromosomal syndromes* (ed. J.J. Yunis), p. 364. Academic Press, New York.

Lichtenstein, J.R., G.R. Martin, L.D. Kohn, P.H. Byers, and V.A. McKusick. 1973. Defect in conversion of procollagen to collagen in a form of Ehlers-Danlos syndrome. *Science* **182**: 298.

Magenis, R.E., F. Hecht, and E.W. Lovrien. 1970. Heritable fragile sites on chromosome 16: Probable localization of haptoglobin locus in man. *Science* **170**: 85.

Maniatis, T., R.C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G.K. Sim, and A. Efstratiadis. 1978. The isola-

tion of structural genes from libraries of eucaryotic DNA. *Cell* **15:** 687.

Maxam, A.M. and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci.* **74:** 1258.

McClelland, A., L.C. Kuhn, and F.H. Ruddle. 1984. The human transferrin receptors gene: Genomic organization, and the complete primary structure of the receptor deduced from a DNA sequence. *Cell* **39:** 267.

McCurdy, P.R. 1971. Use of genetic linkage for the detection of female carriers of hemophilia. *N. Engl. J. Med.* **285:** 218.

McKusick, V.A. 1970. Prospects for progress. *Excerpta Med. Int. Congr. Ser.* **3:** 407.

———. 1971. *Mendelian inheritance in man: Catalogs of autosomal dominant, autosomal recessive, and X-linked phenotypes.* 3rd edition. Johns Hopkins University Press, Baltimore.

———. 1980. The anatomy of the human genome. *J. Hered.* **71:** 370.

———. 1982a. The human gene map. *Clin. Genet.* **22:** 359.

———. 1982b. The human genome through the eyes of a clinical geneticist. *Cytogenet. Cell Genet.* **32:** 7.

———. 1986a. The morbid anatomy of the human genome: A review of gene mapping in clinical medicine (first of four parts). *Medicine* **65:** 1.

———. 1986b. *Mendelian inheritance in man: Catalogs of autosomal dominant, autosomal recessive, and X-linked phenotypes,* 7th edition. Johns Hopkins University Press, Baltimore.

McKusick, V.A. and F.H. Ruddle. 1977. The status of the gene map of the human chromosomes. *Science* **396:** 390.

Miller, D.A., O.J. Miller, V.G. Dev, L. Medrano, and H. Green. 1974. Human chromosome 19 carries a poliovirus receptor gene. *Cell* **1:** 167.

Mirsky, A.E. and H. Ris. 1951. The desoxyribonucleic acid content of animal cells and its evolutionary significance. *J. Gen. Physiol.* **34:** 251.

Mitelmann, F. 1983. Catalogue of chromosome aberrations in cancer. *Cytogenet. Cell Genet.* **36:** 1.

Nadeau, J.H. and B.A. Taylor. 1984. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc. Natl. Acad. Sci.* **81:** 814.

Naylor, S.L., P.A. Lalley, R.W. Elliott, J.A. Brown, and T.B. Shows. 1980. Evidence for homologous regions of human chromosome 3 and mouse chromosome 9 predicts location of human genes. *Am. J. Hum. Genet.* **32:** 158A (Abstr.)

Ohno, S. 1973. Ancient linkage groups and frozen accidents. *Nature* **244:** 259.

Old, J.M., H. Ayyub, W.G. Wood, J.B. Clegg, and D.J. Weatherall. 1982. Linkage analysis of nondeletion hereditary persistance of fetal hemoglobin. *Science* **215:** 981.

Ott, J. 1974. Estimation of the recombination fraction in human pedigrees: Efficient computation of the likelihood for human linkage studies. *Am. J. Hum. Genet.* **26:** 588.

———. 1976. A computer program for linkage analysis of general human pedigrees. *Am. J. Hum. Genet.* **28:** 528.

———. 1985. *Analysis of human genetic linkage.* Johns Hopkins Press, Baltimore.

Patil, S.R., S. Merrick, and H.A. Lubs. 1971. Identification of each human chromosome with a modified Giemsa stain. *Science* **173:** 821.

Patterson, D., C. Jones, H. Morse, P. Rumsby, Y. Miller, and R. Davis. 1983. Structural gene coding for multifunctional protein carrying oratate phosphoribosyltransferase and OMP decarboxylase activity is located on long arm of human chromosome 3. *Somatic Cell Genet.* **9:** 359.

Phillips, J.A., III, B.L. Hjelle, P.H. Seeburg, and M. Zachmann. 1981. Molecular basis for familial isolated growth hormone deficiency. *Proc. Natl. Acad. Sci.* **78:** 6372.

Reeders, S.T., M.H. Breuning, K.E. Davies, R.D. Nicholls, A.P. Jarman, D.R. Higgs, P.C. Pearson, and D.J. Weatherall. 1985. A highly polymorphic DNA marker linked to adult polycystic kidney diease on chromosome 16. *Nature* **317:** 542.

Robson, E.B., P.E. Polani, S.J. Dart, P.A. Jacobs, and J.H. Renwick. 1969. Probable assignment of the α locus of haptoglobin to chromosome 16 in man. *Nature* **223:** 1163.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:** 5463.

Schnedl, W. 1971. Analysis of the human karyotype using a reassociation technique. *Chromosoma* **34:** 448.

Schoen, R.C., H.C. Summers, and R.P. Wagner. 1984. Thymidine-kinase activity of cultured cells from individuals with inherited galactokinase deficiency. *Am. J. Hum. Genet.* **36:** 815.

Schwartz, D.C. and C.R. Cantor. 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37:** 67.

Schrott, H.G., L. Karp, and G.S. Omenn. 1973. Prenatal prediction in myotonic dystrophy: Guidelines for genetic counseling. *Clin. Genet.* **4:** 38.

Seabright, M. 1971. A rapid banding technique for human chromosomes. *Lancet* **II:** 971.

Slightom, J.L., A.E. Blechl, and O. Smithies. 1980. Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21:** 627.

Smith, C.L. and C.R. Cantor. 1986. Pulsed-field gel electrophoresis of large DNA molecules. *Nature* **319:** 701.

Solomon, E. and W.F. Bodmer. 1979. Evolution of sickle varient gene. (Letter). *Lancet* **I:** 923.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503.

Steinmann, B., L. Tuderman, L. Peltonen, G.R. Martin, V.A. McKusick, and D.J. Prockop. 1980. Evidence for a structural mutation of procollagen type I in a patient with the Ehlers-Danlos syndrome type VII. *J. Biol. Chem.* **255:** 8887.

Su, T.-S., R.L. Nussbaum, S. Airhart, D.H. Ledbetter, T. Mohandas, W.E. O'Brien, and A.L. Beaudet. 1984. Human chromosomal assignment for 14 argininosuccinate synthetase pseudogenes: Cloned DNAs as reagents for cytogenetic analysis. *Am. J. Hum. Genet.* **36:** 954.

Sumner, A.T., H.J. Evans, and R.A. Buckland. 1971. New technique for distinguishing between human chromosomes. *Nature* **232:** 31.

Szabo, P. and D.C. Ward. 1982. What's new with hybridization *in situ? Trends Biochem. Sci.* **7:** 425.

Tonegawa, S. 1985. The molecules of the immune system. *Sci. Am.* **253:** 122.

Watson, J.D. 1976. *Molecular biology of the gene,* 3rd edition, p. 428. W.A. Benjamin, Menlo Park, California.

Watson, J.D. and J. Tooze. 1981. *The DNA story: A documentary history of gene cloning.* W.H. Freeman, San Francisco.

Watson, J.D., J. Tooze, and D.T. Kurtz. 1983. *Recombinant DNA: A short course.* Scientific American, New York.

Weiss, M. and H. Green. 1967. Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc. Natl. Acad. Sci.* **58:** 1104.

White, R., M. Lippert, D.T. Bishop, D. Barker, J. Berkowitz, C. Brown, P. Callahan, T. Holmes, and L. Jerominski. 1985. Construction of linkage maps with DNA markers for human chromosomes. *Nature* **313:** 101.

Wilson, E.B. 1911. The sex chromosomes. *Arch. Mikrosk. Anat. Entwicklungsmech.* **77:** 249.

Wyman, A.R. and R.L. White. 1980. A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci.* **77:** 6754.

Yang, F., J.B. Lum, J.R. McGill, C.M. Moore, S.L. Naylor, P.H. van Bragt, W.D. Baldwin, and B.H. Bowman. 1984. Human transferrin: cDNA charcterization and chromosomal localization. *Proc. Natl. Acad. Sci.* **81:** 2752.

Young, B.D., M.A. Ferguson-Smith, R. Sillar, and E. Boyd. 1981. High-resolution analysis of human peripheral lymphocyte chromosomes by flow cytometry. *Proc. Natl. Acad. Sci.* **78:** 7727.

Yunis, J.J. 1976. High resolution of human chromosomes. *Science* **191:** 1268.

————. 1983. The chromosomal basis for human neoplasia. *Science* **221:** 227.

Yunis, J.J., J.R. Sawyer, and K. Dunham. 1980. The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. *Science* **208:** 1145.